

NIM – ERCAP PDF

ERCAP Request #81686 for FY2007

1. Principal Investigator

Name	Phone	Email
Bethel, Edward (Wes)	510-486-7353	ewbethel@lbl.gov

2. Other Authorized Preparers

3. Senior Investigators	(Utah) Chris Johnson, Chuck Hansen, Steve Parker (LLNL) Valerio Pascucci, Hank Childs (ORNL) Sean Ahern (UCD) Ken Joy
-------------------------	--

4. Project Title	SciDAC2 Visualization and Analytics Center for Enabling Technologies
------------------	--

5. Project Name	VACET
-----------------	-------

6. Project Class	SciDAC
------------------	--------

7. Sponsoring Site	USA: Lawrence Berkeley National Laboratory
--------------------	--

8. Science Category	Computer Sciences
---------------------	-------------------

9. DOE Office and Program	ASCR – Computer Sciences
---------------------------	--------------------------

10. Is this project funded by the DOE Office of Science?

X

Yes, this project has direct grant support from the DOE Office of Science

Who is your DOE Program Manager?

Yukiko Sekine

DOE Office of Science Grant Number(s)

1. SciDAC2 Visualization and Analytics Center for Enabling Technologies (VACET). DOE Grant number N/A as of 10/24/06. Total award amount is \$2.2M/yr for five years.

2. High Performance Visualization, MICS Base Program. Award is \$500K yr through FY09.

According to LBL Finance people, there are no "Grant Numbers" associated with either of these awards but both are "on the books."

NIM – ERCAP PDF

11. Does this project make use of national security information?

Yes

☒ No

If this project is given a NERSC award, I agree to monitor the usage associated with it to ensure that, to the best of my ability to determine, usage is for the project described here.

☒ Yes

No

For continuing projects: I have audited the MPP and/or HPSS usage associated with this project, and to the best of my ability to determine, all usage was for the project specified.

Yes

No

12. Computational Resources Requested

Center	Resource	Alloc Type	Repo	Hours Used 2006	Hours Requested 2007
NERSC	MPP (<u>Usage Charging</u>)	DOE Production	m636		20,000

13. Mass Storage Resources Requested

Center	Resource	Alloc Type	Repo	SRUs Used 2006	SRUs Requested 2007
NERSC	<u>HPSS</u>	DOE Production	m636		250,000

14. Justification for Resources Requested

SRU's: During the first year, we are estimating the need to store up to 50TB of data. The data will come from our science stakeholders in accelerator, astrophysics, combustion, climate, fusion; not all science stakeholders will produce this data at NERSC. The 50TB request is only an estimate, but is conservative based upon customer projections for data size/complexity in the coming year.

MPP hours: The 20K hours figure is based upon an estimate of time needed to perform visualization and analytics processing on 50TB worth of data in a prototype/test/revise cycle over the course of one year. We expect our technologies and use patterns span the range from large, single-process tasks that consume vast amounts of memory through modestly parallel algorithms (up to 256 CPUs). We anticipate most use to be interactive rather than batch. During the first year of our work with NERSC, we'll establish a baseline for estimating out-year MPP requests based upon first-year usage and attenuated with next-year projections.

15. PDSF Repos

16. Project Description

16.1 Project Summary: Provide a brief project description. What will this project accomplish? What is the significance of this work?

Launched in 2006 as one of nine centers under the Department of Energy's Scientific Discovery through Advanced Computing (SciDAC-2), VACET focuses on leveraging scientific visualization and analytics software technology as an enabling technology for increasing scientific productivity and insight. Advances in computational technology have resulted in an information big bang, which in turn has created a significant data understanding challenge. This challenge is widely acknowledged to be one of the primary bottlenecks in contemporary science. The vision for our Center is to respond directly to that challenge by adapting, extending, creating when necessary and deploying visualization and data understanding technologies for our science stakeholders at DOE's open computing facilities (NERSC/LBNL and NCCS/ORNL). Using an organizational model as a Visualization and Analytics Center for Enabling Technologies (VACET), we are well positioned to be responsive to the needs of a diverse set of scientific stakeholders, including other SciDAC projects, in a coordinated fashion using a range of visualization, mathematics, statistics, computer and computational science and data management technologies.

More specifically, VACET will provide visualization and analytics software infrastructure to support the challenging data understanding needs of SciDAC Science Applications. This infrastructure will be deployed at DOE's open computing facilities, both NERSC/LBNL and NCCS/ORNL.

16.2 URL for a relevant web page.

<http://www.vacet.org/>

17. Accomplishments

The project is new in FY07.

18. Relevant Publications

18.1 Refereed Publications: List all refereed publications in the last 12 months based on research using NERSC resources. You may include publications submitted to journals but not publications in preparation.

18.2 Other Publications: List up to 5 other relevant publications in the last 12 months, also based on research using NERSC resources.

19.1 Code and Application Descriptions

Code Name	Description	Machines	Algorithms	Languages	Libraries	Other Software
VisIt	VisIt is a parallel-capable visualization application the uses a client-server architecture for displaying	Davinci (70%) Bassi (10%) Jacquard (10%) NERSC-5 (5%)	Visualization/analytics	c++, MPI, MPI/IO	hdf4, hdf5, netcdf, OpenGL, Xlib, Qt, Posix communication	

results of visualization to a remote client.		Seaborg (5%)	&I/O, Boxlib, Chombo	
Code Name	Checkpoint	Performance Limits	Performance Comments	Planned Code Enhancements
VisIt	N	MPI I/O, Posix I/O	As a visualization application, the absence of graphics hardware at NERSC will be a major concern in terms of a performance bottleneck.	Over the course of the project, VACET will be adding capabilities to VisIt to meet the needs of our science stakeholders.

19.2 Code and Application Performance

20. Storage Required on Computational Systems

Home: 100GB per team member, estimate up to 10 possible team members having NERSC accounts.
Scratch needed for largest job: 25TB in year 1.

21. Data Intensive / HPSS Project Requirements

In the first year, our most likely use pattern is to move data from HPSS to scratch, then perform vis/analytics processing reading data from scratch. Visualization results, which will most likely consist of images, will be sent to the remote user. Possible delivery mechanism include remote GLX, VirtualGL, Chromium Renderserver (which uses VNC's RFB protocol to communicate with a remote VNC Viewer application via a VNC Proxy that runs at NERSC in user space).

In later years, we anticipate collaborating with Shoshani's SDM Center and NERSC Analytics SDM specialists to avoid, where feasible, the step where we manually move data from mass to secondary storage. An SRM or SRB-like layer, combined with data subsetting that would "run on HPSS", would potentially reduce the amount of data movement between HPSS and scratch storage.

In terms of I/O requirements, our project will likely be one of the most data intensive at NERSC in loading data to vis/analytics applications from secondary storage. We expect 1TB datasets to be routine in FY07, and at 1GB/s I/O to/from scratch, it will take 17 minutes to load 1TB worth of data (and about 291 hours to load 1PB of data — given DOE's emphasis on Petascale computing, visualization, etc., we are hoping for an across-the-board improvement in I/O bandwidth at DOE's computing facilities). We feel a more aggressive posture with regard to I/O would be helpful for our project, and would suggest as some intermediate milestones: 10GB/sec in FY07, 100GB/sec in FY09, 1TB/sec in FY11.

22. Networking

23. Other HPC Support

A "startup-type" of allocation from NCCS/ORNL to provide support to stakeholders who compute/vis/analyze there.

24. Additional Information

It is important to be mindful of the fact that the VACET's activities complement NERSC's existing investment in Visualization/Analytics as well as DOE's investment in basic/applied visualization research. VACET will serve as a "bridge" between research and production visualization and analytics. It acts as a consumer of visualization and analytics research. Its deliverables will be deployed for production use at DOE's production computing facilities by both scientific users as well as production visualization programs. The VACET team has representatives from both NERSC Analytics and the equivalent at ORNL; we (and the program office) felt it important that any visualization effort in SciDAC have representation from visualization efforts at DOE's open computing facilities.

We listed VisIt under the "code" section as it is a known quantity. It is most likely we will be running a "custom" version of VisIt, as opposed to the one in production use already at NERSC. There will likely be other visualization and analytics "codes" we will run, some reasonably well known (e.g., SCIRun) while others will be research prototypes (a.out's). They will read and process all manner of data produced by stakeholders in accelerator, astro, climate, combustion, and fusion. All these codes don't fit easily within the categories in this ERCAP request, which offers a categorization system most appropriate for numerical rather than vis or analysis codes that focus on data-intensive activities.

25. Feedback and Project Requirements

To date, David Skinner has been very accommodating in his role as SciDAC Outreach Center PI in terms of soliciting input from our team. While we understand the mission scope for the Outreach Center is undergoing refinement and clarification, David helped set our team up with virtual web hosting on rohan, where www.vacet.org is now on the air. We are very grateful for his help in that regard.

We are hopeful that David's vision will come to fruition with the deployment of the GForge infrastructure at NERSC. VACET can definitely benefit from its features: role-based access to content/documents, charting/graphing for project management, SVN & CVS access to software, tracking software downloads, etc.

We look forward to a fruitful collaborative relationship with the NERSC program and the SciDAC Outreach Center.

26. Current Request Status

Not Finished: This request is not yet ready for review.

X

Finished: This request is finished and ready for review.